DOCUMENT RESUME

ED 054 229 24 TM 000 825

AUTHOR Borg, Walter R.
TITLE Three Levels of Evaluation for Educational Products.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-6-2931
PUB DATE 3 May 71
CONTRACT OEC-4-7-062931-3064
NOTE 10p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Behavior Change, Classroom Observation Techniques, Educational Methods, Educational Objectives, Educational Planning, *Educational Programs, Evaluation, *Evaluation Techniques, Experience, *Models, Performance, Program Costs, *Program Evaluation, Validity

ABSTRACT

Three levels of evaluation that can be used in the assessment of educational products and processes are: 1) Unvalidated Form of Experience, 2) Validated Form of Experience, 3) Direct Performance Evaluation. Each of these evaluation models is described in detail, and factors involved in selection of the evaluation model are discussed. (MS)

ED0 54229

BR 6-2931
PA 24

5/3/71

EA

THREE LEVELS OF EVALUATION FOR EDUCATIONAL PRODUCTS

by Walter R. Borg

The purpose of this paper is to describe briefly three levels of evaluation that can be used in the assessment of educational products and processes such as those developed by the Far West Laboratory. The ideas expressed in this paper have evolved gradually over a period of several months as a result of discussions in the weekly Program Directors' meeting at the Laboratory. In all likelihood, by the time this paper is discussed at the Executive Panel Meeting, the ideas will have evolved still further. However, it seems desirable at this point to restate some of our views about evaluation within the context of these three levels. It should be recognized that evaluation strategies can be divided into many more than three levels. However, the levels discussed here appear to be fairly basic and are viewed as a worthwhile point of departure for considering the whole question of product - process evaluation. We have labeled the three evaluations models: (1) Unvalidated Form of Experience, (2) Validated Form of Experience, and (3) Direct Performance Evaluation.

Unvalidated Form of Experience

In efforts that approach the development task at the unvalidated form of experience level, the investigator first hypothesizes that certain kinds of experience should bring about the changes in pupil behavior which are his ultimate objective. He then develops a product or process that is designed to provide these designated experiences to the child.[1] In

1. In this paper, it is assumed that we are concerned with products that have as their ultimate objective changing the performance of children.

TM 000 825

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

evaluating, he collects observational data to determine whether the kinds of experiences which he attempted to build into his product do, in fact, occur in situations where the product is used. For example, let us suppose that the investigator hypothesizes that carrying out specified play activities with wooden blocks of different size and shape will increase the child's ability to make size and shape discriminations when presented with objects other than wooden blocks. To evaluate his product in the classroom environment using the unvalidated form of experience approach, he would observe in the classroom to determine whether children do, in fact, have wooden blocks of different size and shape available and carry out the specified play activities with these wooden blocks. If he found that the blocks were available and the children did play with them, he would conclude in his evaluation that the desired form of experience was provided by the product, and would infer that children exposed to this experience would increase in their ability to discriminate size and shape. Thus, the unvalidated form of experience approach to evaluation requires the investigator to make a rather large inferential leap. Specifically, he must infer without supporting evidence that his original hypothesis is valid, i.e., that there is a relationship between the form of experience and the performance outcomes of the learner.

Form of experience, although the lowest level of evaluation described in this paper, is still one step beyond the process that has been used in the development of a great many educational products. The process that the typical author employs in building an educational product (such as a new curriculum) is to assemble a product which, in his opinion, will provide certain experiences to the learner and these experiences, in turn,

will result in bringing about desired behavioral changes on the part of the learner. This approach involves a much greater inferential leap than the unvalidated form of experience approach, since the author is assuming not only that the experience will lead to performance changes, he is also assuming that the materials he has developed will create a learning environment in which the desired experiences actually take place. This assumption is rarely tested by observing use of the product in learning situations. Essentially, the author is basing his evaluation on face validity, i.e., the product appears to be suitable for the purpose intended. This approach relies entirely on examination of the product per se, while unvalidated form of experience is based on evaluation of the product in use.

Validated Form of Experience

Evaluation which is based on validated form of experience again involves determining if the product provides the forms of experience intended by the investigator. However, this approach contains an additional element which reduces the inferential leap that must be made by the investigator. This additional element is related research evidence. Again, the investigator hypothesizes that certain experiences will lead to certain performance changes on the part of the children who are exposed to his product. However, his hypothesis in this case is supported by research evidence which shows a relationship between the kind of experience he intends to provide and the kind of performance change he wishes to bring about. This evidence will not have been collected using the product the investigator has developed. Some of the evidence will probably have been collected by other investigators using products that are similar to all or part of the product being

evaluated. The investigator may also choose to supplement this outside supporting evidence by carrying out carefully controlled small scale laboratory studies in which the relationship between some aspect of the experience his product provides and the desired change in pupil performance is investigated. In many ways, the validated form of experience approach to product evaluation is analogous to the construct validity approach to test development. In establishing the construct validity of a test, the test developer carries out studies which evaluate the performance of the test against the relevent theoretical constructs that form the foundation for the test. To return to our "blocks" example, the investigator would be conducting a validated form of experience evaluation if he were able to report studies in which it was found that children who manipulated objects of different sizes and shapes (such as blocks) showed improved size and shape discrimination when tested with other objects (such as toy automobiles and trucks). The inferential leap that the investigator must make in this case is that he must assume that research evidence related to the concept or theoretical construct which forms the rationale for his product, will hold for the specific product that he is developing. Or to give a hypothetical example, he must conclude that because Jones found in 1965 that chimpanzees who were given wooden cut-outs of different sized circles, and crescents to play with could better discriminate between large and small bananas, organes and tangerines; children given similar size and shape discrimination experiences will similarly transfer their learning to different objects.

Direct Performance Evaluation

In the third form of evaluation, the investigator sets up an experimental design in which a group of children is randomly divided and either exposed or not exposed to the product or process being developed. After this exposure, the performance of the children who are exposed to the product is compared with the performance of the comparable control group on the specific product objectives.[2] While the unvalidated and validated form of experience evaluations approach the question of pupil performance indirectly, this third form of evaluation is concerned directly with that pupil performance which occurs as a result of exposure to the specific product being developed by the investigator. To return to our size and shape discrimination example, if the investigator wished to conduct a direct performance evaluation of his process, he would assign pupils randomly to a treatment and control group and then expose treatment group pupils to the prescribed experiences involving manipulation of wooden blocks of different size and shape. At the end of this exposure, he would test both the treatment and the control group pupils on a series of size and shape discrimination problems involving objects other than wooden blocks. If he found that pupils in the treatment group made significant gains in their ability to discriminate objects in the test exercises and that no gains were made by pupils in the control group, he could conclude that his product had been successful in changing the performance of children in the treatment group. Even at this level of evaluation, an inferential

---

2. When working with variables such that there is virtually no chance of performance change on the part of the control group, the investigator may choose to employ a single group design with pre-post evaluation. This approach reduces cost and increases the risk of drawing invalid conclusions.

leap is still required if the investigator proposes that his product be used with some broad population of children similar to those in his treatment group. He must assume that children in his treatment group are a representative sample of the broadly defined population that constitutes his target for the product.[3] The only way such a conclusion can be drawn with a specified degree of confidence is by randomly selecting the treatment group from the population.[4] For most broadly defined populations (for example, all pre-school children in the United States between the ages of 3 and 5), it is extremely difficult and costly to work with random samples. Remember, that for a sample to be random, every individual within the defined population must have had an equal chance of being selected as a member of the treatment group. Even if we abandon a simple random sample and move to a three or four stage process for obtaining the random sample, the logistics are extremely difficult and studies carried out on such samples are likely to be expensive since we can logically expect individuals in the sample to be widely dispersed. Of course, the investigator can define a more narrow population such as "all pre-school children in the city of Berkeley between the ages of 3 and 5," and then draw his samples from this population. This does not resolve his dilemma, however, since now his results will only apply directly to the Berkeley population. Most educational product developers ignore this dilemma by seeking a sample with no obvious bias and making the aforementioned inferential leap.

---

3. This assumption must also be made with the two forms of experience models unless subjects are selected randomly from some relevant population.
4. "Specified degree of confidence" referes to statistical confidence intervals.

Selecting the Evaluation Level

Selecting a level of evaluation depends on a number of factors. Let us review four of the most important factors that must be considered in deciding what level of evaluation will be employed with a given product.

Nature of the Problem

Perhaps the most important factor is the nature of the pupil performance that must be evaluated in order to achieve the product objectives. Some types of pupil performance, such as skill in number operations, can reasonably be expected to show up shortly after exposure to an effective product. Others, such as improved self-concept, are more likely to require a lengthy exposure to the product or process before any measurable changes are likely to occur. Even if such changes do occur, the state of the art in measuring such variables as self-concept is far behind our ability to measure simpler performance variables such as accuracy in addition. Thus, the nature of the variables with which we are working often rules out direct product evaluation because of time lag, measurement difficulties, or other such problems.

Funds Available

For most educational products, any competent evaluation specialist can outline procedures for the direct evaluation of performance that would yield results that could be accepted with a high degree of confidence. However, in most areas such ideal evaluation procedures would be so costly as to be unthinkable from a practical viewpoint. Therefore, in selecting an evaluation procedure, the investigator must consider the different costs involved in reaching different levels of confidence. He must then make

some compromise between the cost incurred and the benefits obtained in terms of validation data, etc. The cost-benefit approach, for example, rules out the use of random sampling in most educational product evaluation studies.

Other Constraints

In addition to money, there are other considerations that sometimes limit the investigator in his choice of an evaluation approach. Perhaps the most obvious of these constraints is the time limit. If a direct evaluation of pupil performance would require three years while a series of small-scale studies that would permit a validated form of experience evaluation could be done in a year, it may well be that the investigator must accept the one-year approach regardless of his preferences. If, for example, he only has funds available for a one-year evaluation and has reason to believe that additional funds will not be forthcoming, it would be foolish to launch an evaluation effort that would require three years to complete.

Consequences of Being Wrong

A factor that must be considered in the investigator's decision to carry out evaluation at a particular level is the consequences of drawing conclusions from evaluation which may later prove to be invalid. For example, let us suppose that an investigator obtains favorable results from a series of studies which comprise a validated form of experience evaluation and later discovers in a direct performance evaluation of the product that his original conclusions were incorrect. What effect would this initial mistake have on the children who have been exposed to the

product during the interim? If there is likely to be little or no negative effect, then the investigator is justified in using the least expensive evaluation approach which meets his minimum requirements. However, if a mistake can cause serious undesirable consequences, he is morally bound to use evaluation techniques which will permit him to draw conclusions about the performance outcomes of his product with a high degree of confidence. Fortunately, the consequences of being wrong in the development of most educational products are not as serious as might be the case in areas such as medicine. The reason that the consequences of being wrong are relatively less serious in education is that the changes in pupil performance brought about by a new product, no matter how ineffective the new product may be, are probably not going to be significantly worse than what is happening to the child already.

What Can We Conclude About These Three Levels of Evaluation?

It is possible to draw certain conclusions about the levels of evaluation we have discussed. First and perhaps most important is that, although direct performance evaluation of the product is intrinsically superior to validated form of experience which, in turn, is intrinsically superior to unvalidated form of experience, there are situations in which each of these evaluation approaches is appropriate. Furthermore, when dealing with a product or process that seeks to make a number of performance changes, the investigator may find it necessary to evaluate different objectives using different levels of evaluation. For example, in evaluating Minicourse 1, all three evaluation levels were used. For variables such as prompting, no evidence of a relationship with pupil performance was available and none was collected in the course evaluation.

Data on the teachers' use of prompting, however, were collected which indicated that a form of experience (i.e., teacher prompts) had been present in the learner's environment. Thus, prompting was subjected to an unvalidated form of experience evaluation.

In the case of higher order questions, research data from previous studies has shown a relationship between teacher use of higher order questions and pupil performance. Our evaluation showed that teachers increased their proportion of such questions after taking Minicourse 1. Thus, higher order questions were evaluated at the validated form of experience level. For some objectives, such as increasing the amount of pupil participation, we made direct measures of pupil performance before and after teachers had taken the course. These measures indicated that pupils of teachers who had taken Minicourse 1 participated significantly more in discussion lessons. Thus, amount of pupil participation was evaluated at the direct performance level.